Classification of Chemicals Based on Structured Toxicity Information

Matthew T. Martin

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Environmental Sciences and Engineering.

Chapel Hill
2008

Approved by:

David J. Dix, Ph.D
Ivan Rusyn, M.D. Ph.D.
Richard Judson, Ph.D.
Donald Fox, Ph.D.

ABSTRACT

Matthew T. Martin

Classification of Chemicals Based on Structured Toxicity Information

"Under the direction of Drs. David J. Dix and Ivan Rusyn"

Thirty years and millions of dollars worth of pesticide registration toxicity studies, historically stored as hardcopy and scanned documents, have been digitized into highly standardized and structured toxicity data within the Toxicity Reference Database (ToxRefDB). Toxicity-based classifications of chemicals were performed as a model application of ToxRefDB. These endpoints will ultimately provide the anchoring toxicity information for the development of predictive models and biological signatures utilizing *in vitro* assay data. Utilizing query and structured data mining approaches, toxicity profiles were uniformly generated for greater than 300 chemicals. Based on observation rate, species concordance and regulatory relevance, individual and aggregated effects have been selected to classify the chemicals providing a set of predictable endpoints. ToxRefDB exhibits the utility of transforming unstructured toxicity data into structured data and, furthermore, into computable outputs, and serves as a model for applying such data to address modern toxicological problems.

ACKOWLEDGEMENTS

The research described here has been partially supported by the U.S. Environmental Protection Agency. This report has been reviewed by the U.S. Environmental Protection Agency's Office of Research and Development. The contents do not necessarily reflect the views and policies of the Agency nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ACToR – Aggregated Chemical Toxicity Resource

CEBS - Chemical Effects in Biological Systems

DER – Data Evaluation Record

DSSTox – Distributed Structure-Searchable Toxicity

FFDCA – Federal Food, Drug and Cosmetic Act

FIFRA – Federal Insecticide, Fungicide and Rodenticide Act

FQPA – Food Quality and Protection Act

HTS – High-throughput screening

IRIS – Integrated Risk Information System

LOAEL – Lowest Observed Adverse Effect Level

LOEL – Lowest Observed Effect Level

MOA – Mode-of-Action

MOU – Memorandum of Understanding

NAS – National Academy of Sciences

NCGC – NIH Chemical Genomics Center

NIH – National Institutes of Health

NIEHS – National Institute of Environmental Health Sciences

NRC – National Research Council

NOAEL – No Observed Adverse Effect Level

NTP – National Toxicology Program

OECD – Organization for Economic Co-operation and Development

OPP – Office of Pesticide Programs

OPPTS – Office of Prevention, Pesticides and Toxic Substances

RED – Reregistration Eligibility Decision

SAR – Structure-Activity Relationship

ToxRefDB – Toxicity Reference Database

U.S. EPA – United States Environmental Protection Agency

Chapter 1

Literature Review

The United States Environmental Protection Agency (EPA) has identified roughly 9,000 environmental chemicals that have been or may need to be assessed for their human exposure and toxicity potential (Judson et al In Press). In order to fully assess the toxicity of these chemicals, the current testing paradigm requires *in vivo* mammalian toxicity studies that require thousands of animals, millions of dollars and years to complete. In an effort to investigate the utility of alternative toxicity testing strategies, three large-scale efforts have begun to test libraries of chemicals, including pharmaceuticals, industrial chemicals and pesticides. The National Institutes of Health (NIH) Chemical Genomic Center (NCGC), National Toxicology Program (NTP) and EPA each have research programs for generating *in vitro* assays results on hundreds if not thousands of chemicals (Collins et al 2008; Inglese et al 2007; NTP 2004; Dix et al 2007). Current efforts in QSAR (quantitative structure-activity relationship) model development are working through chemical space and into biological activity and *in vitro* assay results to bolster predictive power (Zhu et al 2008).

The combination of unprecedented amounts of biological data being generated and the development of methods for integrating and analyzing these diverse datasets makes it imperative that the anchoring endpoints and adverse outcomes be just as computable and

biologically-relevant. A common thread among these developing technologies and approaches is the need for reference toxicity information and detailed toxicity classifications of chemicals.

The amount of reference toxicity information on environmental chemicals, including primary studies, study reviews and summarized reports, quickly diminishes beyond pesticide active ingredients, Integrated Risk Information System (IRIS) chemicals, NTP nominated chemicals and a few other sources (Judson et al In Press). There is little direct literature on detailed toxicity-based chemical classification. However, efforts to digitize and structure the vast stores of open literature and unpublished industry-submitted studies will provide the information in a context amenable to classifying chemicals with respect to their toxicity. The currently available chemical-induced toxicity databases vary widely in breadth and depth of information (Yang et al 2006a, 2006b; Bitsch et al 2006). The Yang et al papers (2006a, 2006b) summarize the available toxicity databases and places them into various categories based on their content and structure. IRIS is a good example of a database that has large content, covering greater than 500 chemicals and multiple toxicities, but lacks the standardization and detailed relational structure to provide accurate and efficient read-across (U.S. EPA 1997). Yang et al further describe systems that store literature citations and summary toxicity information such as TOXNET, which are invaluable resources for chemical-specific literature searches and safety assessment, but lack searchability and read-across. ToxML, and related Food and Drug Administration (FDA) databases, and REPDOSE are two examples of relational formats that currently house hundreds of chemicals and multiple study types in a standardized format, including controlled vocabularies (Yang et al 2006a; Bitsch et al 2006). These databases primarily cover

pharmaceuticals and industrial chemicals, but are limited in their coverage of pesticides. There is no direct literature on the robust toxicity dataset produced for pesticide active-ingredients, but through regulatory mandates many pesticides undergo a full suite of mammalian toxicity testing.

In addition to the toxicity database efforts, more chemical-centric databases and curation efforts have begun to identify the landscape of toxicity information associated with environmental chemicals (Judson et al In Press). The EPA DSSTox program is dedicated to hand curated chemical structure and using the chemical structure as a link to external data sources (US EPA 2007; Richard 2004), which provides an invaluable resource for aggregating information across varying domains of information. These chemical information domains are well characterized in Judson et al (In Press), and are broken down into chemical structure, physicochemical properties, biochemical assay data, *in vivo* toxicology assay data primary tabular and secondary tabular, *in vivo* toxicology test reports via URL, *in vivo* toxicology summary calls, regulatory listings, chemical categories and phenotypes. The information is stored in the EPA ACToR (Aggregated Computational Toxicology Resource) database, which also uses chemical identity and structure as a primary link between data sources (Judson et al 2008). Many of the data sources within ACToR were culled from the internet.

Web accessible toxicological data sources have been previously characterized (Felsot 2002; Russom 2002; Junghans et al 2002; Winter 2002; Wolfgang 2002; Young 2002; Patterson 2002). These internet resources range from food and drug toxicity to environmental and ecological toxicity. Some of the internet sources provide fairly detailed summaries from cancer-related and genotoxicity studies. However, the information from

these various sources is dispersed across the internet and in a wide variety of formats. Systems such as TOXNET, DSSTox, ACToR and PUBCHEM have made many of these resources available in a compiled format able to be searched based on chemical structure.

With existing efforts to make available, digitize and structure the toxicity information landscape for environmental chemicals, researchers have begun to compute with the compiled information for a variety of purposes. Analysis of legacy toxicity data for understanding the importance of specific toxicity tests and their role in the risk assessment process is underway. Reproductive toxicity study retrospective analyses have sought to understand the role of the second generation in hazard identification and the overall assessment of reproductive toxicity (Janer et al 2007a). Additionally, retrospective analyses on developmental toxicity studies measuring the value of running a second species through developmental toxicity studies. (Janer et al 2007b). These retrospective efforts demonstrate the ability to take structured toxicity information and test hypothesis through data analysis.

Using legacy toxicity information for analyzing species concordance has assisted in risk assessment decisions for specific tumor types and assisted in rodent to human extrapolation. Gold et al (2001) characterized species concordance for 1458 chemicals that did or did not cause tumors in various species. A similar approach was taken in the pharmaceutical industry, but with a focus on the concordance between human and animal toxicities with an overall conclusion supporting animal testing (Olson et al 2000).

Additionally, efforts to use surrogate or shorter-term *in vivo* endpoints to predict long-term outcomes have demonstrated the use of legacy toxicity information for predictive toxicology. In Mathews et al (2005), gene mutation in Salmonella and *in* vivo micronucleus genetic toxicity studies showed good correlation for predicting carcinogenicity. In Allen et al

(2004), specific shorter-term liver pathologies were used as forecasters of liver tumor formation. These studies may be limited in their application for risk assessment or other regulatory toxicology applications, but are quality examples of utilizing legacy toxicity data in a computable manner across a relatively large set of chemicals.

The next step taken, as shown in the Zhu et al (2008) paper, was the incorporation of screening data or alternative testing data, including genomics, into predictive toxicology. In Fielden et al (2002), approaches for predicting toxicity using *in silico* methods and alternative testing data was laid out for toxicologist to advance the understanding of the molecular basis of toxicity. Iconix's Drug Matrix® stored experimental information from genomic studies including detailed pathology and developed genomic signatures or classifiers predictive of toxicity (Fielden et al 2005; Fielden et al 2007) and showed promise in predicting toxicities of environmental chemicals (Martin et al 2007). Importantly, the use of reference toxicity information was used in the development of the classifiers in all studies.

Similar governmental efforts to create the data management tools for storing genomic and phenotypic information has created the computational environments for the analysis of large genomic datasets with corresponding toxicity or phenotypic data. NIEHS's Chemical Effects in Biological Systems (CEBS) has been developed to store diverse biological information resulting from various toxicity and biological studies (Waters et al 2003). Systems-based toxicology in the world of drug discovery and drug safety assessment has begun to take hold and used as a viable approach both early in the discovery process and later in assessing toxicological information (Mayne et al 2006). Some of the tools that are making this possible include ingenuity pathway analysis, and in Fliri et al (2005) the ingenuity pathway analysis tools along with other analyses demonstrated that linking *in vitro* assay

results to drug label information and adverse effect data provided mechanistic insight into purported toxicities and side-effects of drugs. Similar system-based and pathway-based approaches for toxicity prediction to limit the high attrition rate of pharmaceuticals in the pipeline have produced other tools and products (Apic et al 2005). These analytical tools required extensive curation of the biological literature and resulted in large databases for storing the information.

The developmental of biological databases, including the controlled vocabularies that enable read-across, have pushed forward the field of toxicology, but have also identified data gaps in both general toxicity information and the molecular basis for the toxicity. Furthermore, novel challenges have arisen in the field of toxicology and more broadly the field of biology due to the ever increasing complexity and size of generated datasets and the need for standardization across those datasets, whether it is traditional toxicology, genomics or screening data. These challenges have given rise to the field of bioinformatics to assist in solving the issues and challenges by integrating computer science with biological sciences (Roos 2001). The field of toxicology continues to utilize bioinformatics tools and resources, but there are emerging needs for further database and analytical tool development, including the digitization of legacy pesticide toxicity information into relational databases to make the information accessible to the scientific community.

The development of reference toxicity and detailed pathway and cellular network databases can provide the context for interpreting generated molecular- and cellular-level data. The reference toxicity information can be thought of as a form of phenotypic anchoring even though the information is extrinsic to the experiments. The importance and role of phenotypic anchoring for *in vivo* toxicogenomics studies has been well laid out (Paules

2003). The fundamental principle of phenotypic anchoring has experimentally shown to provide clearer profiles of biological perturbation. In Powell et al (2006), toxicity endpoints and protein adduct formation was used to phenotypically anchor oxidative stress gene expression due to acetaminophen exposures. There are many other examples of using the concept of phenotypic anchoring for deriving differentially expressed genes and genomic classifiers predictive of the final endpoint (Fielden et al 2005). With high-throughput and high-content assays being available direct phenotypic anchoring is not possible in the same way as *in vivo* experiments. External sources of phenotypic anchoring, including reference toxicity information, are needed to provide the context for the *in vitro* experiments. In general, the data generation and analysis as applied to toxicology has integrated a broad set of scientific disciplines and has formed a sub-discipline called computational toxicology. The breadth of research in the field of computational toxicology was outlined in Kavlock et al (2007) and further demonstrates the need for data generation and analysis consistency.

Chapter 2

Introduction

In an order to progress toward alternative toxicity testing and novel predictive methods as laid out by the National Academy Sciences' (NAS) National Research Council (NRC) in "Toxicity Testing in the Twenty-first Century: A Vision and a Strategy" (NRC 2007), the scientific community must recycle existing legacy data, through digitization, in order to further enable the driving technologies. Historically, the traditional toxicity studies performed by industry in support of pesticide registration have been used for the development of risk assessments on a single compound or representative group of compounds and can cost up to $10 million dollars per chemical. This vast store of high quality guideline legacy toxicity information on hundreds of compounds has thus far been electronically inaccessible. The electronic capture and structuring of pesticide toxicity information alone will serve as an invaluable resource for both retrospective and prospective scientific efforts.

Although an extensive body of open literature toxicity studies is available, the ability to automate data mining of unstructured information and extract uniform toxicity endpoints across a large chemical set has not been demonstrated. Initiatives to electronically store the vast amounts of legacy toxicity data into databases has been characterized previously (Yang

et al 2006a, 2006b). A portion of these efforts have successfully stored toxicity study information at varying levels of granularity in a relational format utilizing an XML standard (e.g., ToxML), controlled vocabularies, and/or standardized data models (e.g., REPDOSE (Bitsch et al 2006)). The chemical coverage for these databases includes pharmaceuticals and industrial chemicals, but is limited in their coverage of pesticides.

Pesticide manufacturers undergoing registration and reregistration of pesticide products and formulations through the EPA are mandated under the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) to meet specific data requirements, one of which is toxicological testing. There are various levels of toxicological testing required based on use pattern, production volume and other factors. All existing pesticides active ingredients registered before November 1, 1984 must be reevaluated for their effects on human health and the environment, due to various legislative mandates including the 1988 FIFRA amendments and the FIFRA and Federal Food, Drug and Cosmetic Act (FFDCA) as amended by the Food Quality Protection Act of 1996 (FQPA). New pesticides active ingredients, meaning any ingredients introduced since 1978, have required extensive testing to progress from development to registration. Specifically, food-use pesticide active ingredients require a complete set of *in vivo* mammalian oral toxicity studies due to human oral exposure potential. The results of EPA's review on a chemical's product chemistry, efficacy, toxicology, environmental fate and effects and exposure assessment are primarily summarized in Reregistration Eligibility Decision (RED) documents.

In order to complete the RED documents and other similar reviews, registrant-submitted toxicity studies are reviewed by the agency for data quality and scientific content in Data Evaluation Records (DERs). Each DER includes reviews on individual studies for

their adherence to Office of Prevention, Pesticides and Toxic Substances (OPPTS), OPPTS/Office of Pesticide Programs (OPP) and Organization for Economic Co-operation and Development (OECD) health effect guidelines that have been established in various forms over the years. DERs also supply detailed study design, categorical endpoint, critical effect and complete dose-response information.

The EPA and other regulatory agencies are investigating novel approaches to predict toxicity in order, for instance, to reduce the number of animals required for toxicity testing, to increase mechanistic understanding of chemical toxicity and to carry out large scale screening of chemicals that have not previously been fully characterized. All of these efforts require a body of high quality *in vivo* toxicity data in order to test and validate new approaches. To assist these efforts, the EPA is developing a searchable compilation of data from regulatory studies and compiling this data into a database called ToxRefDB (Toxicological Reference Database). The ToxRefDB effort is initially focused on entering subchronic rodent, developmental rat and rabbit, multigeneration reproduction rat, chronic/cancer rat and cancer mouse studies. The database schema is generalized to capture all OPPTS, OPP, OECD mammalian toxicity guideline studies on technical grade chemicals, including additional study types such as 28-day neurotoxicity and developmental neurotoxicity studies. Additionally, detailed taxonomical effect vocabularies have been developed for repeat measure effects such as clinical chemistry, hematology and urinalysis, for terminal target organ observations such as organ weight, gross pathology and non-neoplastic and neoplastic pathology, and for non-organ-directed toxicity such as clinical signs, neurotoxicity, developmental toxicity and reproductive toxicity.

An important initial application of ToxRefDB is to provide anchoring *in vivo* toxicity data for the EPA ToxCast™ research program, which has been designed to address the agency's needs for chemical prioritization by using state-of-the-art approaches in high-throughput screening (HTS) and toxicogenomics (Dix et al 2007). Nearly all of the ToxCast Phase I chemicals are food-use pesticide active ingredients and have undergone the full suite of mammalian toxicity tests making for an unparalleled reference set of toxicological information. The complete and highly standardized dataset provided by ToxRefDB facilitates analysis of the ToxCast Phase I chemicals across chemical, study type, species, target organ and effect.

Finally, ToxRefDB serves as a model for other efforts to capture quantitative, tabular toxicology data from legacy and new studies, and to make this data useable for cross-chemical computational toxicology analysis.

Chapter 3

Methods

Data Characteristics

The reviews on the registrant-submitted toxicity studies, known as Data Evaluation Records or DERs, were collected for roughly 400 chemicals. The file types of the DERs include TIFF, Microsoft Word, Word Perfect and PDF formats, some of which are not text-readable. Every DER file was then indexed based on a file name convention that consisted of the OPP Pesticide Chemical Code (PC Code), study identification number (MRID), study type identification number (based on 870 series OPPTS harmonized health effect guidelines), species code, review identification number (TXR) and a review version code, which identified the review as a primary review, secondary review, supplemental review, updated executive summary, or a deficient review. In total 4,620 DERs were indexed spanning roughly 3,000 studies. The searchable file structure created an efficient work-flow for database population. Each study assesses a single technical grade chemical's toxicity potential in a single species, spanning developmental, reproduction, subchronic, chronic and cancer toxicities. DER formats have changed over time, but the underlying content has remained consistent. The first portion of the DER outlines the test substance, purity, lot/batch numbers, MRID, citation, OPPTS guideline and reviewers of the study. The

executive summary captures all of the basic study design information, including species and strain, doses, number of animals per treatment group and any deficiencies in study protocol. In addition, the executive summary describes the most relevant observed effects and establishes the appropriate No Observed Adverse Effect Level (NOAEL) and Lowest Observed Adverse Effect Level (LOAEL) endpoints for the study based on the identified critical effects. The next sections, which are test material and animal information, can be used to verify the test substance identity and purity and to provide detailed species/strain and husbandry information. Full dose response information is then provided in text and tables under a variety of headings, which in this thesis will be referred to as 'effect type' and listed in order of appearance within most DERs. For most DERs these are mortality, clinical signs, clinical chemistry, hematology, urinalysis, gross pathology, non-neoplastic pathology and neoplastic pathology. For reproductive and developmental studies parental, offspring, reproductive, maternal and fetal effects are listed separately. Within each effect type heading, 'effect target' (i.e., clinical parameters or target organs) results are displayed. Some effect targets can be described simply as increasing or decreasing, whereas pathological results are presented as specific 'effect descriptions', e.g., hypertrophy and hyperplasia.

Relational Model

In the development of ToxRefDB, a relational model approach was taken with input from other subject-specific database model approaches, including ToxML. The resulting data model is therefore semi-hierarchical in nature: a single compound was tested in multiple studies; each study contained multiple treatment groups; multiple effects could be observed in each treatment group. The data model was conceptualized from a chemical-centric view to propagate data integration and exchange across various systems and to facilitate linking of

the reference toxicity information to chemical-specific data generated using *in vitro* technologies. Simplifying constraints, based on OECD/OPPTS harmonized health effect guidelines, identify study design parameters that must be met, ranging from the purity and administration methods of the test compound to the number of animals in a treatment group. The relational model was then implemented into a table structure with established relationships ensuring data integrity, updateability and standardization (Figure 1). Specific components of the relational data model are highlighted individually in Appendix A, B, C and D.

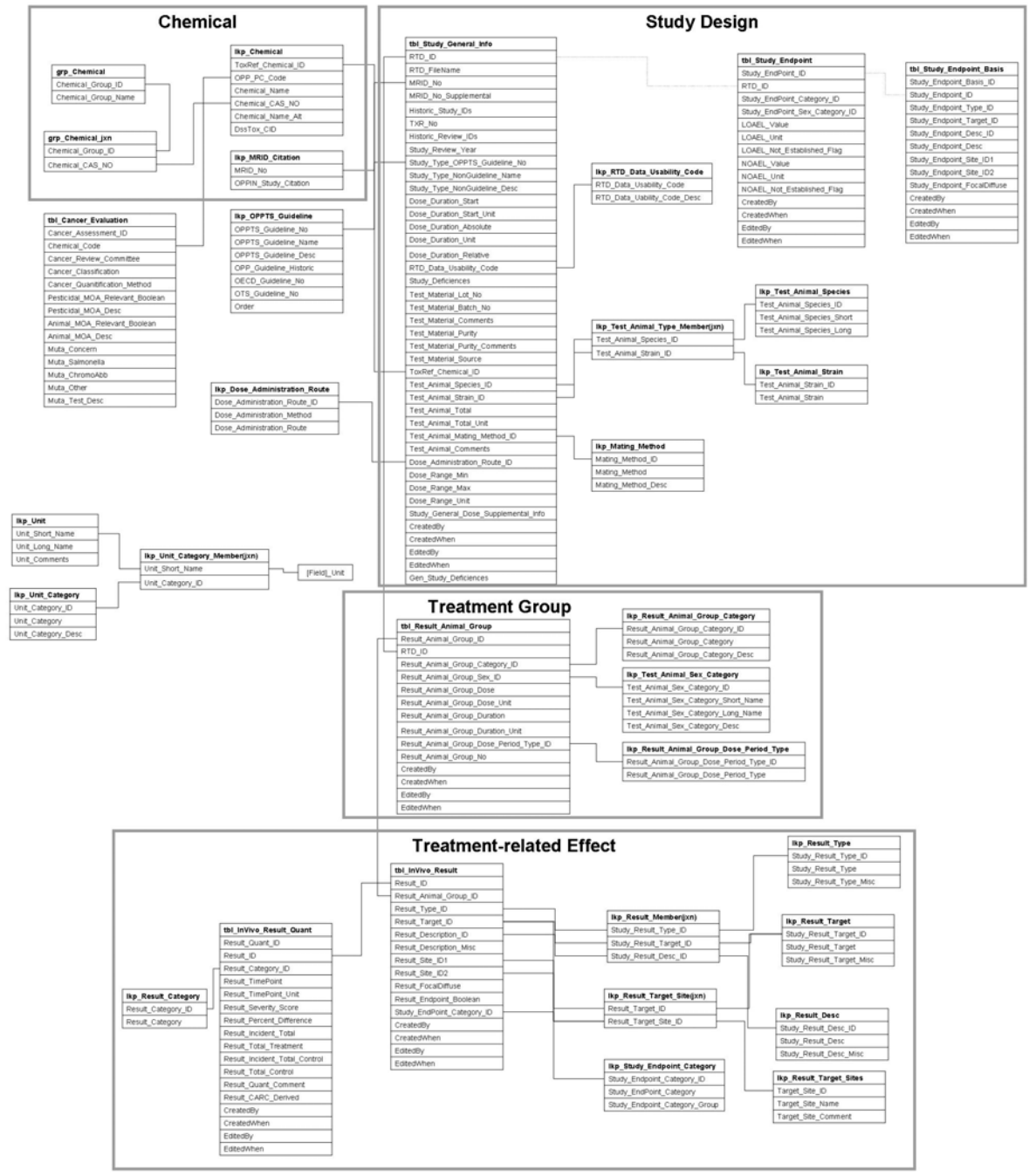


Figure 1. ToxRefDB Relational Data Model

Development of a Toxicity-based Controlled Vocabulary

The development of controlled vocabularies within ToxRefDB was necessary for the standardization of data captured across the various study types and studies performed over

roughly 30 years. The non-redundant list of terms across various domains provided data integrity and searchability.

The chemical information within ToxRefDB has relied on the chemical identification and structural curation within EPA's DSSTox Program (U.S. EPA 2007) and the chemical data management within ACToR (Aggregated Computational Toxicology Resource) (Judson et al In Press). ACToR will link the toxicology data in ToxRefDB to the high-throughput screening (HTS) data being generated through the ToxCast program.

The study type vocabulary was based on the unique study types harmonized by OECD and OPPTS (U.S. EPA 1996). Specific standardized terminology for study design was established for species/strain, method/route of administration and units for dose and duration. Treatment group-related vocabularies were developed to establish the generation, gender, and dosing period.

A primary goal in reviewing the registrant-submitted toxicity studies is to establish NOAEL/LOAEL pairs for a variety of categorical endpoints, including systemic, offspring, maternal, parental, developmental and reproductive toxicity, all across the study types. These categorical endpoints are captured and normalized across studies at the effect level, enabling a direct link to the critical effects in which the NOAEL/LOAEL was derived.

The development of a toxicological effect vocabulary was approached in a domain-specific manner, with clinical pathology terms being derived from existing literature, reproductive and developmental toxicity terminology collected from various collaborative resources including the International Life Sciences Institute's (ILSI) Developmental Toxicology Working Group and organ pathology terms collated from the National Toxicology Program's (NTP) "Pathology Code Tables" (NTP 2007). The vocabulary then

underwent further standardization by mapping all synonymous terms to a single non-redundant value. A taxonomical approach was then taken for establishing the finalized effect vocabulary based on a three-tiered hierarchical model with the effect type being the top layer, followed by effect target and by effect description. Examples of effect type include clinical chemistry, hematology, urinalysis, body weight, mortality, gross pathology, non-neoplastic pathology, neoplastic pathology, developmental and reproductive effects. Subclasses of these types include specific target organs (e.g., liver, lung, spleen, etc.) or measured analytes (e.g., ALT, AST, cholesterol, etc.). The specific combinations of effect type and target are then further sub-classed based on a non-redundant descriptive term (e.g., increase, decrease, hypertrophy, atrophy, etc.). Specific to the organ pathology terms, each target organ has a set of regions, zones and cell types that characterize the site of toxicity. A common representation of the data throughout the manuscript relies on the hierarchical nature of the vocabulary and will be represented as such with study type at the highest level, then tested species, followed by the combination of effect type, target and description. Vocabularies were developed under a standardized and taxonomic approach. Further groupings and relationships between entities have been established that begin the development of a toxicity endpoint-based ontology.

Data Input

The ToxRefDB Data Entry Tool was developed in Microsoft Access® and provides the user interface for all initial data input. Following initial quality control, discussed below, the data is migrated to ToxRefDB, which has been implemented using the open source MySQL™ platform. The ability to utilize the legacy toxicity data entered into ToxRefDB requires consistent and accurate data entry. The initial phase of data entry has consisted of a

series of protocols, outlined in a ToxRefDB Standard Operating Procedures (SOP) document, that call for mapping the toxicological information to standardized fields and vocabulary and extracting treatment-related effects from any given study. Data entry priority has been broken down by study type, with the subchronic rodent, chronic/cancer rat and cancer mouse studies being entered first, followed by multigenerational rat studies and developmental rat and rabbit studies. The next phase of data entry will involve the entry of additional study types in collaboration with OPP following completion of the initial dataset.

Data Quality Control and Management

Entered studies have undergone up to 100% cross-checking, which entails having secondary data entry personnel validate each entered value based on the source information (primarily the DERs). Internal quality control (QC) consists of continued cross-checking of studies by data entry personnel, systematic updates of ToxRefDB to ensure consistency across the studies and a tiered QC approach for the entered studies. The tiered approach involves up to 10% independent QC. Error rates greater than 2% trigger 100% QC of related fields or records.

Data Output and Analysis

Once quality control procedures have been conducted, analytic methods can be applied to specific ToxRefDB outputs. In order to ensure consistency and repeatability of analysis a data format output template was established and directly queried using the ToxRefAnalysis program, which is written in Java™. The first column consists of concatenated chemical information including CAS registry number and chemical name. The second column represents the effect or endpoint and is implemented primarily as a concatenated set of fields representing study type, species and the effect (combination effect

type, target and description) or aggregated effects (effect group name). The final column is primarily the lowest observed effect level (LOEL), but can be categorical or Boolean outputs as well. ToxRefAnalysis cross-tabulates the result set from ToxRefDB and can perform specific data manipulation functions, including distinguishing missing study results from negative results and filtering out effects or chemicals that do or do not meet specific requirements. The resulting dataset is a matrix of chemicals in the first column and effects along the first row, with LOEL filled in where appropriate. The format is highly amenable to statistical data analysis, including descriptive and predictive data mining algorithms.

In order to assess statistically significant species concordance across different effects, a permutation study was carried out. For each effect, the association between chemical and effect for the rat and mouse study was randomly permuted one thousand times. The cross-species concordance for all simulations (permutations) was recorded and compared to the observed concordance, thus giving an estimate of the concordance due purely to chance. Analyses were carried out using R version 2.6.1 (Ihaka and Gentleman 1996).

Chapter 4

Results

Summary Data Characterization

ToxRefDB captured *in vivo* mammalian toxicity study information from DER spanning 411 pesticide active chemicals. A subset of these chemicals is being used in the first phase of the ToxCast program. The focus of this thesis was on the entire set of ToxRefDB chemicals; however the resulting toxicity-based classifications of chemicals have been applied to the ToxCast chemical set. Furthermore, this thesis focuses on systemic toxicity and cancer endpoints culled from subchronic rat, chronic/cancer rat and cancer mouse studies, which cover 334 chemicals.

ToxRefDB enabled analysis to be performed along toxicologically relevant axes, including by chemical, NOAEL/LOAEL, categorical endpoint, effect, aggregated effect group, study type and species. Study duration, dosing methods, data quality, guideline adherence and gender were additional parameters for filtering or analyses. Initial analysis was performed to assess regulatory relevance, commonality across chemicals, consistency across study types and species concordance. By looking across all chronic/cancer rat, cancer mouse and subchronic rat studies, 31,427 effects were assigned to 4,431 different treatment groups in a total of 831 studies (Table 1).

	Chemicals	Studies	Treatment Groups	Treatment Groups w/ Effects	Effects[a]	Critical Effects[b]
Total	***334***	***831***	***9,466***	***4,431***	***31,427***	***4,865***
Subchronic Rat	236	251	2,179	1,370	11,796	1,739
Chronic/Cancer Rat	281	300	4,228	1,721	12,215	1,822
Cancer Mouse	266	280	3,059	1,340	7,416	1,304

(a) - Total number of effect type, target, and description combinations assigned to any treatment group

(b) - Effects that are criteria for establishing the systemic LOAEL

Table 1. ToxRefDB Summary Statistics

With individual effects being represented as a combination of study type, species, effect type, effect target and effect description, analyses at varying levels of this effect taxonomy focused downstream analysis. Of the 31,427 effects 1,287 unique effects were observed, of which 601 were deemed critical effects in at least a single study. In order to begin to characterize the chemicals based on these effects, the distribution of effects by effect type enabled comparisons and honed in on the most relevant classes of effects (Figure 2). The distribution of critical effects revealed that non-neoplastic pathologies predominate systemic endpoint selection based on the high percentage of NOAELs/LOAELs driven by non-neoplastic pathology. This demonstrates the regulatory relevance of this class of endpoints. Treatment-related changes in body weight also contribute significantly to systemic endpoint criteria. However, systemic LOAELs were established based solely on body weight changes in 73 studies primarily at the high dose or maximum tolerated dose (MTD). Observation rates were similar across study type and species with the exception of the clinical chemistry, hematology and neoplastic pathology, which were not routinely assessed due to study design or guideline requirements. Therefore, study design constraints limited the ability to provide cross-species or cross-study classifications for clinical chemistry, hematology and neoplastic pathology. Additional factors, including high rates of body weight changes and corresponding organ weight changes were consequences of study

design due to guideline requirements of testing up to the MTD in the chronic/cancer studies and using the subchronic study to establish the MTD. Of the chemicals that caused neoplastic lesions in the rat or mouse chronic/cancer studies, 35% caused neoplastic lesions in both rat and mouse. We define the percentage of chemicals that cause an effect in both rat and mouse over the total that cause the effect in only the rat or mouse the "species concordance" for that endpoint. Species concordance for non-neoplastic pathology was 68% and consistency between the subchronic and chronic/cancer rat study was 74%.

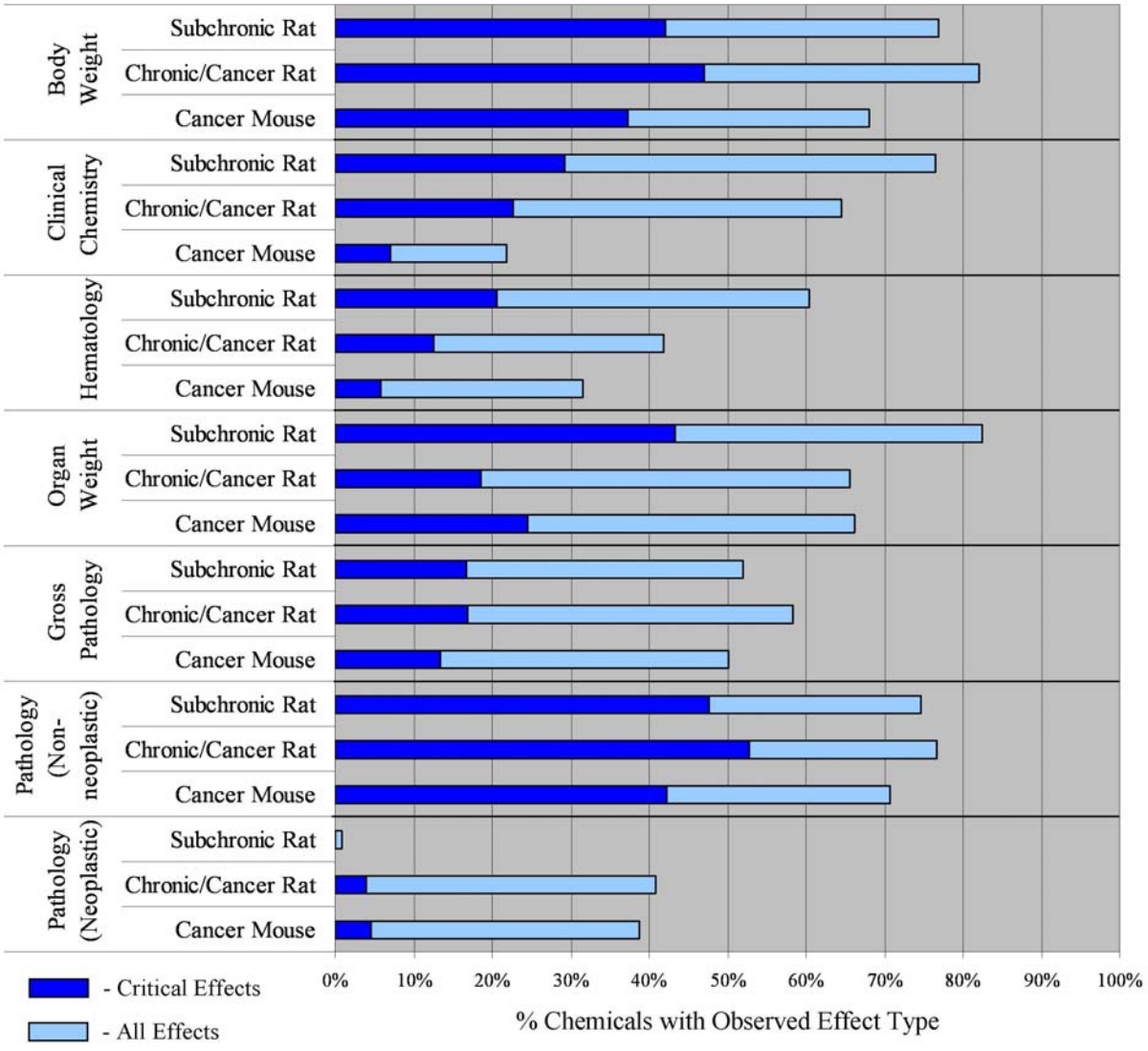


Figure 2. Distribution of Effects by Type

Observation rate analysis of 334 chemicals across cancer mouse, chronic/cancer rat and subchronic rat studies. % Observed is represented by the colored bars and is calculated as the percentage of chemicals with the observed effect type across the three studies. The dark blue bars indicate the critical effects that derived the systemic NOAEL/LOAEL endpoint, whereas the light blue represents all other effects. Non-neoplastic pathology critical effects are observed at the highest rate across all three study types.

Non-neoplastic pathology drove systemic endpoint selection, i.e., NOAEL/LOAEL levels, while neoplastic pathology results inherently inform regulatory cancer classification. The distribution of these pathological responses for the same 334 chemicals across target organ characterized the regulatory relevance, observation rate and identified organs that were further investigated for specific pathological effects (Figure 3). Greater than 50% of the chemicals caused a treatment-related pathological response in the liver and greater than 30% in the rat kidney. This observation made these organs obvious toxicologically-relevant targets and prime candidates for exploring individual effects. Target organ pathology observation rates were similar across study type and species and only the liver and kidney effects were conserved across species as a pathological target at greater than 30%, a rate comparable to neoplastic lesions across all targets.

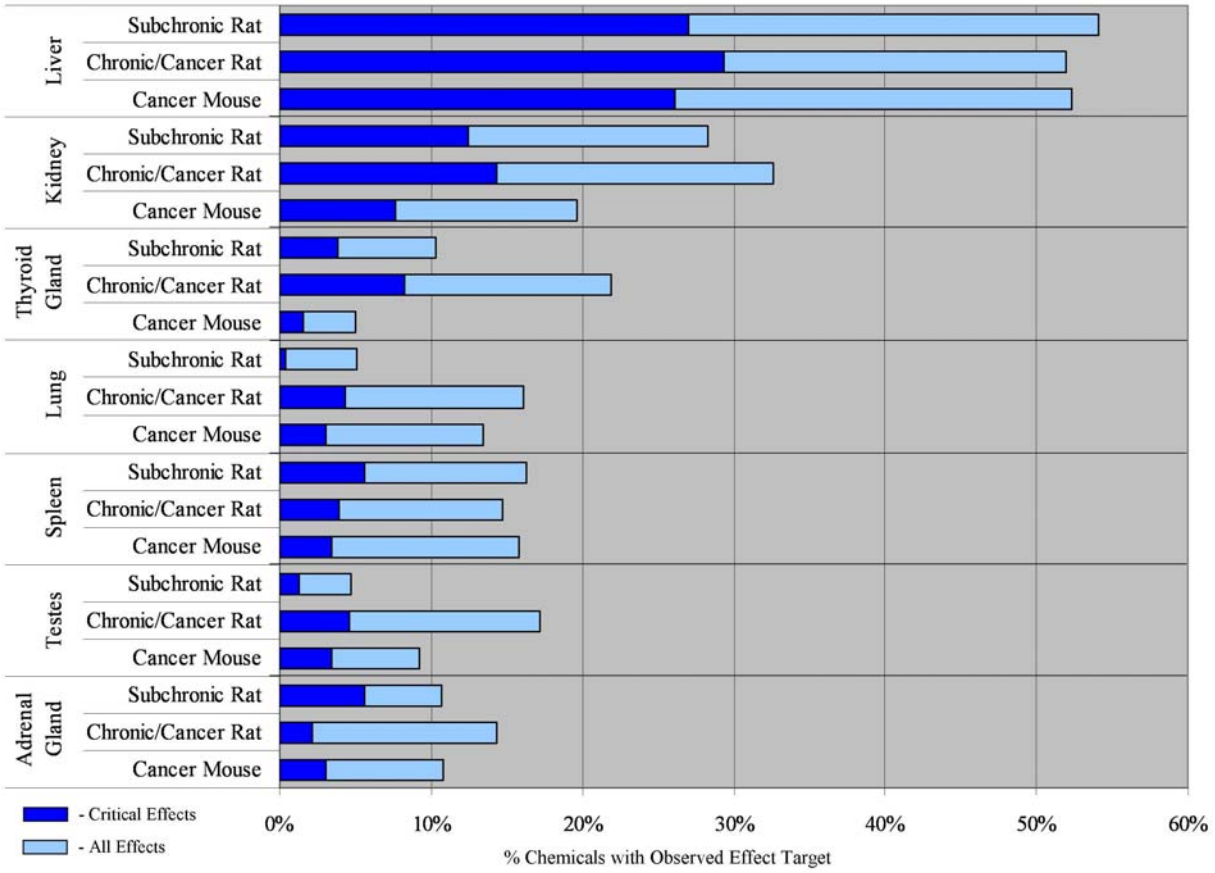


Figure 3. Distribution of Effects by Type
Target organ pathology observation rate analysis of 334 chemicals across cancer mouse, chronic/cancer rat and subchronic rat studies. % Observed is represented by the colored bars and is calculated as the percentage of chemicals with the observed pathology across various target organs. The dark blue bars indicate the critical effects that derived the systemic NOAEL/LOAEL endpoint, whereas the light blue represents all other effects. Liver and kidney pathology effects are observed at the highest rate and are among the most prevalent and sensitive targets for establishing endpoints.

Specific, individual effect descriptions that relate to highly detailed pathological outcomes would provide classifications with the highest biological specificity. Limitations of classifying chemicals based solely on specific individual effects was apparent from our data as evidenced by there being only 12 detailed pathology-related effects that were observed in greater than 10% of the chemicals (Table 2). In addition to low observation rates, biases based on study design and pathology nomenclature limited the overall ability to compare chemical toxicities when individual effects were used. Liver hypertrophy is the

only common effect across both species. Related or near-synonymous terms, such as liver adenoma, combined adenoma/carcinoma and carcinoma, would be more informative if grouped together. In order address the limitations of classifying chemicals based on specific individual effects, biologically-related groupings of effects were derived. Grouping or aggregating effects in a non-arbitrary, biologically-driven manner inherently increased observation rates while maintaining the ability to draw biologically relevant conclusions.

Study Type	Species	Effect Type	Effect Target	Effect Description	% Observed
Chronic	mouse	Pathology (Non-neoplastic)	Liver	Hypertrophy	25%
Chronic	rat	Pathology (Non-neoplastic)	Liver	Hypertrophy	25%
Chronic	mouse	Pathology (Neoplastic)	Liver	Adenoma	21%
Chronic	mouse	Pathology (Non-neoplastic)	Liver	Necrosis	16%
Chronic	mouse	Pathology (Neoplastic)	Liver	Adenoma/Carcinoma Combined	14%
Chronic	rat	Pathology (Non-neoplastic)	Kidney	Nephropathy	14%
Chronic	mouse	Pathology (Non-neoplastic)	Liver	Pigmentation	14%
Chronic	rat	Pathology (Non-neoplastic)	Liver	Vacuolization	12%
Chronic	mouse	Pathology (Neoplastic)	Liver	Carcinoma	11%
Chronic	rat	Pathology (Non-neoplastic)	Thyroid	Hyperplasia	11%
Chronic	rat	Pathology (Neoplastic)	Thyroid	Adenoma	10%
Chronic	rat	Pathology (Non-neoplastic)	Liver	Eosinophilic Focus	10%

Table 2. Individual Pathology Effects Observed in Greater than 10% of Chemicals

Extending Cancer Classification to Proliferative Lesions

Classifying chemicals based on carcinogenic potential is limited to a small set of target organs or broadly classed across target organs as a carcinogen. In order to increase the observation rates across target organ cancer classifications were extended to include all proliferative lesions. In general, only neoplastic lesions are considered indicative of carcinogenic potential, but including non-neoplastic pathologies related to proliferation provides a conservative schema for assessing and predicting carcinogenic potential. For tumor responses, aggregating effects based solely on neoplastic pathology for each target

organ increased classification beyond individual mouse liver tumor effects as shown in Table 2, but remained limited to mouse liver and rat thyroid neoplasia based on an initial >10% observation rate cutoff. Extending the cancer-related classifications to comprise of all proliferative lesions increased the number of target organs classified and included liver, kidney, thyroid, lung and testes. A simulation study was performed to assess if the concordance between rat and mouse occurs at a rate greater than chance across both neoplastic and proliferative classifications (Figure 4). Beyond increasing the overall observation rate, extending chemical cancer classifications to include proliferative lesions significantly increased species concordance.

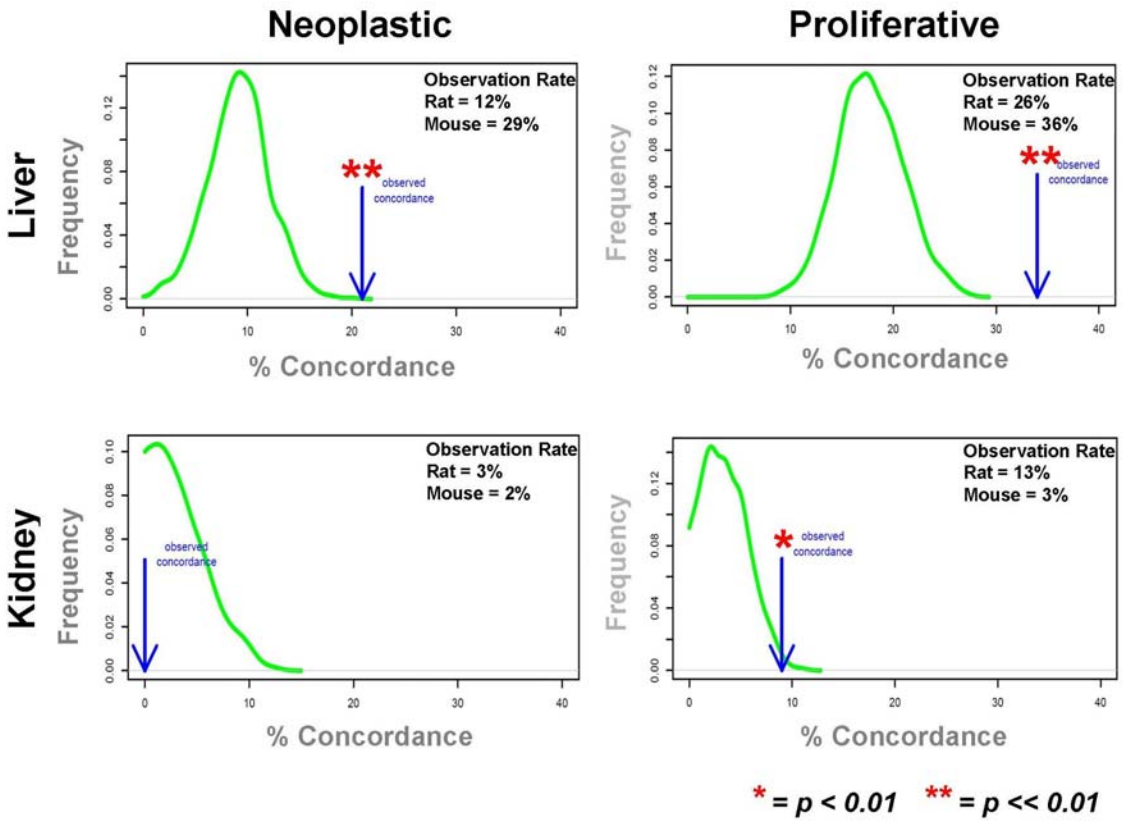


Figure 4. Simulation Study Analysis of Species Concordance
Extending cancer classification to include proliferative lesions increases both observation rate and species concordance. Simulation study using 1000 permutations were performed to compare 279 rat and 260 mouse chronic/cancer study results for neoplastic and proliferative

lesions across the liver and kidney. The simulated empirical density (frequency) based on random alignment of rat and mouse observed rates provides the distribution of concordance expected by chance, whereas the observed concordance is shown with the blue arrow. The single asterisk ($p<0.01$) and double asterisk ($p<0.001$) define statistically significant species concordance.

Toxicity Endpoint and Cancer Progression Schema

Study type, species and dose are examples of attributes and properties that are inherent to the database structure, whereas the relationships between effects are not. Specific biological processes have been well laid out with respect to disease progression, including cancer formation (Hanahan and Weinberg 2000). Although the toxicity data stored in ToxRefDB did not provide molecular insight into cancer progression, effects captured in the database provided key events involved in the progression of a pathological response leading to tumor formation and cancer. Figure 5 conceptualizes the endpoint progression scoring along both an endpoint and cancer progression continuum for which each chemical was assigned.

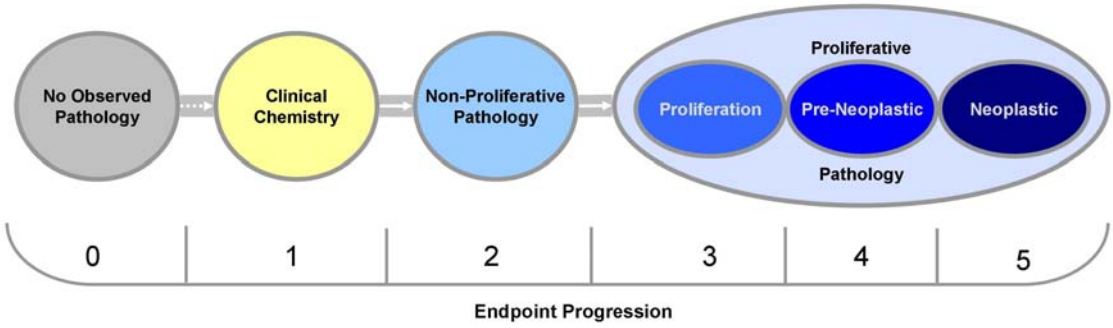


Figure 5. Endpoint Progression Continuum for Ordinal Scoring
Systemic toxicity and cancer progression is illustrated as an endpoint progression continuum. The progression begins with no observed pathology at a given target organ, then to clinical chemistry changes that are pertinent to the target organ, followed by non-neoplastic non-proliferative pathology, including hypertrophy, atrophy, necrosis and inflammation. Progression of these endpoints up to this point is driven by observation rate decline and increasing toxicological relevance. The continuum then progresses toward proliferative lesions and is broken down into three categories: proliferation (cell proliferation and hyperplasia); pre-neoplastic lesions (foci and hyperplastic nodules); neoplastic (tumors).

Endpoint progression scoring reduced the possible classifications from thousands of individual effects to a set of target organs with associated ordinal scores. The distribution of endpoint progression for liver and kidney characterized target level effects without requiring pathology calls along the entire continuum (Figure 6). For example, resmethrin caused treatment-related increases in hyperplastic nodules in the liver, but did not progress to tumor formation. In contrast, metaldehyde caused treatment-related increases in liver tumors, but was not identified as causing any preneoplastic lesions such as hyperplastic nodules or foci, which can be assumed to have occurred as a precursor event to liver tumor formation. Individual or even aggregated effect classifications may miss the associations that these and many other chemicals may have, but endpoint progression scores develop and maintain these associations throughout the analysis. Tumor formation does not necessarily require pre-neoplastic lesions. However, in order to generalize to all target organs and tumor types the order of proliferation to pre-neoplastic to neoplastic was used.

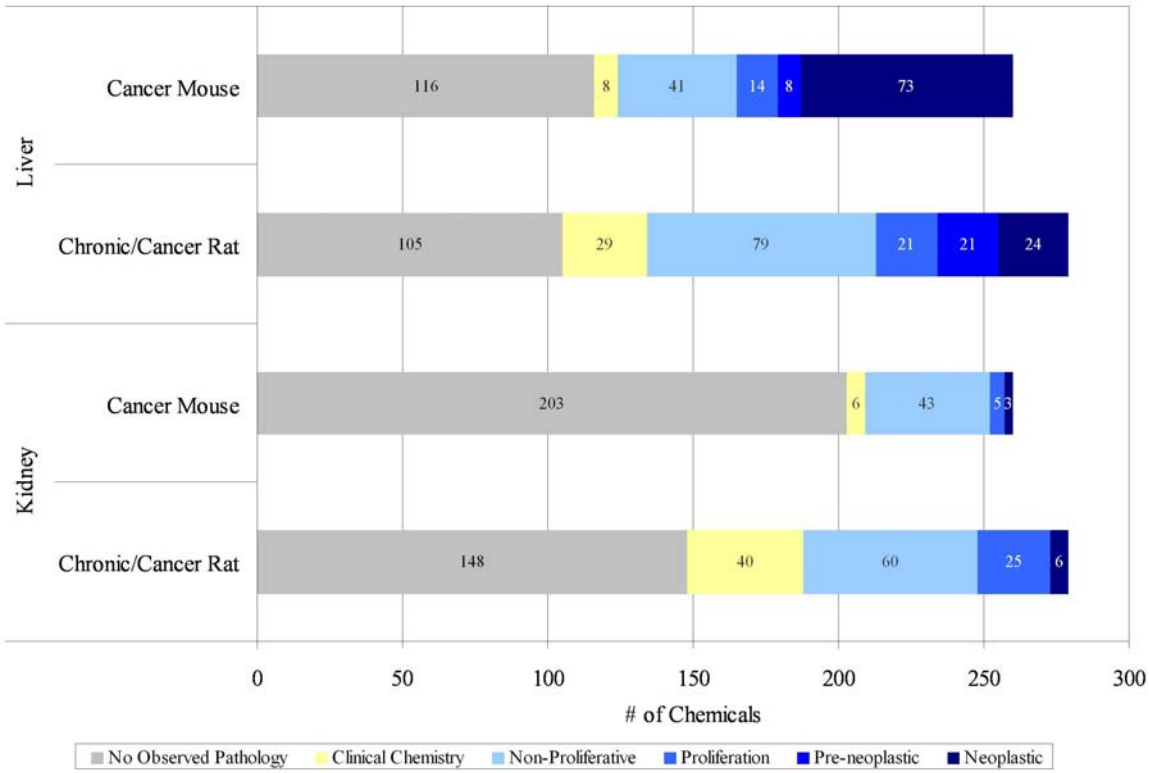


Figure 6. Endpoint Progression Scoring Distribution for Liver and Kidney
Based on endpoint progression, 334 chemicals were scored across 279 rat chronic/cancer and 260 mouse cancer studies for liver and kidney pathology. The chemicals are scored based on the maximum value across the target organ. For instance, if a chemical causes only liver hypertrophy then the chemical would be assigned a 2 for non-neoplastic pathology, whereas if the chemical causes hypertrophy and hyperplasia the chemical would be assigned a 3 for proliferation. Clinical chemistry is target specific, with analytes being labeled by target organ, e.g., ALT for liver and urea nitrogen for kidney.

Potency Ranking

Relative potency across the observed effects provided insight into the sensitivity and relevance of the endpoint and a categorical approach to chemical classification. To derive non-arbitrary dosing intervals, lowest observed doses (mg/kg/day) for body weight changes were analyzed and separated into equivalent quintile bins (data not shown). The resulting bins, ≤15, ≤50, ≤150, ≤500 and >500 mg/kg/day, were then applied to all endpoints. For instance, a chemical that caused liver hypertrophy at 5 mg/kg/day would be assigned a 5, at 25 mg/kg/day a 4 and so on. If the effect was not observed then a zero was assigned.

Potency rankings were used to filter out high-dose effects and to compare across effects. For example, a chemical could be deemed negative for liver weight increase if it was observed at greater than 500 mg/kg/day or if no corresponding liver pathology was observed at or below the observed dose level.

Toxicity-based Classification of Chemicals

With identified target organs, extended effect classes, aggregation of systemic toxicity and cancer effects, toxicity-based endpoints were curated based on toxicological relatedness, biological relevance, observation rate, regulatory relevance and potency. Across rat and mouse studies and various target organs, a subset of endpoints were selected as examples of diverse toxicities and included individual and aggregated effects used for the final classification of the chemical set (Figure 7). Interestingly, the rate of tumorigens and multisite tumorigens were similar between rat and mouse studies, even though target organ-specific tumorigenicity observation rates demonstrated wide variation between species. The selected endpoints along with target organ-specific endpoint progression scores provided uniform characterization of the chemical set at varying levels of toxicological order.

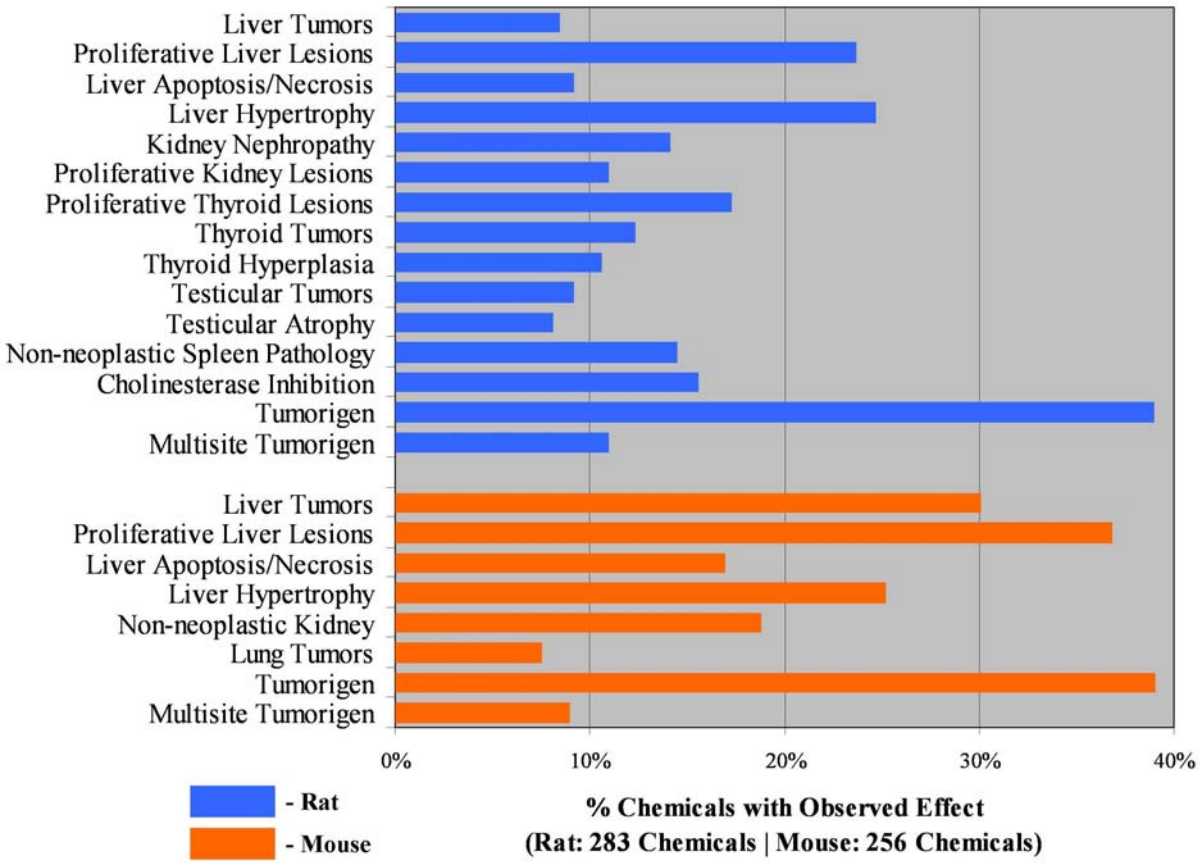


Figure 7. Rat and Mouse Toxicity Endpoints Suitable for Prediction
Selected toxicity endpoints from chronic/cancer rat and cancer mouse studies across 283 and 256 chemicals, respectively, derived from individual and aggregated effects.

Chapter 5

Discussion, Conclusions and Future Directions

Discussion

Chemical toxicity prediction and subsequent validation have not only been limited by the number of input parameters, but by the anchoring endpoints the model or system was developed to replace or predict. Historically, anchoring endpoints and phenotypes have been high-level Boolean classifications, e.g., carcinogen or non-carcinogen (Benigni 1991, Benigni and Zito 2004). Efforts to classify chemicals based on species-specific and target organ phenotypic outcomes from open literature and governmental study reports have required manual collation that is not easily updated and has been limited in endpoint and chemical coverage (Richard and Williams 2003; Richard 2004). Many of the predictive methods assess a chemical's potential to perturb distinct biological processes, whereas these high-level chemical classifications have often been influenced by external factors, including exposure scenarios, risk assessment processes and risk management decisions, thus distancing the classification from the biologically-relevant toxicity potential of a chemical. Ideally, chemical prioritization and risk assessment factors would only subsequently influence the interpretation and application of the predicted outcomes.

Pesticide active ingredients have robust toxicity profiles and are opportune datasets for the advancement of predictive toxicology. By uniformly reviewing toxicity information, DERs present the full-breadth of toxicity information for a single study and summarize treatment-related effects. DERs also contribute to risk assessments and human cancer classifications that begin to identify only the most sensitive or relevant endpoints of concern. Therefore, DERs provide much of the information that goes into chemical safety assessment without losing toxicologically-relevant effects and endpoints screened out in the risk assessment process.

ToxRefDB is the tool for digitizing, storing and structuring the immense amounts of toxicity data in an updateable, searchable and analyzable manner. The development of a standardized vocabulary gave the ability to read-across study types, species and chemicals, thus transforming the manner in which toxicity profiles can be generated. Given a class of compounds or a large set of studies, consistencies and relationships between chemicals or studies can be analyzed in a matter of minutes. In a research application, ToxRefDB can generate toxicity profiles across hundreds of compounds and multiple study types for chemical classification in the ToxCast program.

The framework for reducing greater than 31,000 effects across 334 chemicals into relevant and predictable chemical classifications relied on a combination of measurable factors, including observation or incidence rate, regulatory relevance or critical effect analysis, consistency across study type and species concordance. The effect taxonomy, i.e., study type, species, effect type, target and description, permitted analysis at various levels of granularity. Summary results of effect type and effect target observation rates and regulatory relevance identified pathology, and more specifically liver, kidney, thyroid, lung, adrenal

gland and testis, as predictable endpoints based on an initial 10% observation rate cutoff. The cutoff was established based on an estimated frequency level required for predicting endpoints with high specificity and sensitivity, but was only used as an initial filter. Endpoints of interest, such as rat liver tumorigenicity, were included despite less than 10% observation rate based on regulatory relevance or other factors.

Following effect type and effect target analysis, effect description or individual effect analysis identified a small subset of predictable endpoints, but also demonstrated the limitations of classifying chemicals at such a descriptive level. For some effects or sets of effects it was logical to step up the taxonomy to the effect target level and classify chemicals based on target organ pathology. However, this approach decreases the biological specificity and potentially collapses hundreds of effects into a single endpoint. An example approach of developing biologically-driven groups of effects was the extension of cancer classification beyond tumorigenicity to include all proliferative pathology, which not only increased observation rates but also species concordance. The increased species concordance further demonstrated that shared proliferative responses across species better characterized a chemical's toxicity potential and began to distinguish between species-specific susceptibility for tumor formation and mode-of-action or pharmacokinetic differences. Other factors, such as pathology nomenclature changes over time (Wolf and Mann 2005), may also explain why extending neoplastic lesions to proliferative lesions increases species concordance, but because many of the rat and mouse chronic/cancer studies were run in conjunction with each other nomenclature bias should be minimal. Nonetheless, the approach provides an example of aggregating individual effects to create powerful cancer-related endpoints that would otherwise not be possible for target organs such as the kidney.

Toxicity-based classification of chemicals is limited to a small subset of target organs, including the liver, kidney, thyroid, testis and lung, and may only apply to a single species. The notion of endpoint progression addresses limitations in individual effect classification due to pathology nomenclature, study design, dose spacing and reporting across the above target organs and extends target organ toxicity to an ordinal designation that broadens the target organs classified. Overlaying potency information onto all endpoints provides additional categorical data in the form of dosing intervals facilitating cross-chemical and cross-endpoint comparisons. Potency rankings can also begin to distinguish high-dose and secondary effects versus sensitive target organ-specific effects.

Utilizing the endpoint selection framework, a subset of all observed effects has been identified as anchoring chronic systemic toxicity and cancer endpoints for Phase I of the ToxCast program. A combination of chemical classifications based on individual effects, aggregated effects and organ-level endpoint progression encompass systemic and cancer effects with observation rates that ensure predictability along with biologically-relevant endpoints that enable application and biological verification of generated prediction models.

Conclusions

Unparalleled amounts of legacy toxicity information on pesticides have been captured in a structured format, which provides a platform for repeated and updated chemical characterization and classification. The ability to search and filter across 30 years worth of toxicity data required extensive amounts of data normalization, annotation and curation and was made possible through the development of a robust standardized vocabulary spanning most fields and data elements within ToxRefDB. Application of structured toxicity information to the classification of large chemical sets, e.g., ToxCast Phase I chemicals,

required further data processing using manual and automated structured data mining approaches. Based on specific requirements, including observation rates and regulatory relevance, an endpoint selection framework was applied to the complete dataset and in turn created a manageable set of endpoints for which the chemical set was classified. Whether the analyses of ToxRefDB data represent retrospective, modeling, or research applications, ToxRefDB serves as a resource for scientists, risk assessors and regulators to begin to look across a larger landscape of chemical and toxicity space.

Future Directions

Upon completion of data entry and quality control, similar endpoint selection processes will be applied to the multigeneration reproductive and prenatal developmental study data to determine a set of anchoring endpoints to predict using HTS and genomic data generated through the ToxCast research program. In addition to expanding the toxicity coverage to other study types, ToxCast Phase I non-pesticides or chemicals without DERs will undergo a full literature search and literature review process to fill the data gaps where studies are available. Upon review, the studies will be entered into ToxRefDB. A formal QA/QC process, as described in the methods section, will be performed on all entered studies and will provide the necessary review for eventual public release of the data. A staged public release of the data stored within ToxRefDB is planned following internal EPA review. The initial phase will consist of providing the final outputs, e.g., matrix of chemicals and associated effects and endpoints, directly or through database management systems such as EPA's ACToR. The initial phase is to include the chronic, cancer and reproductive endpoints. The second phase will include the release of developmental toxicity and revisions of previously released endpoints in the final matrix format. The third phase would be the

availability of the entire database for detailed searching, possibly through the development of a web-based query tool.

This novel application of ToxRefDB for the purpose of classifying chemicals demonstrates the ability to transform unstructured information into structured data and to transform structured data into computable data. The next step for the generated toxicity profiles is the anchoring of the endpoints to HTS and genomic data generated within the ToxCast research program. Many machine learning and predictive algorithms will be used along with novel methods that apply to the diverse dataset of biochemical, molecular, and cell-based assay data (Kavlock et al In press). Additionally, the structured information can be reformatted into computable outputs specific to other analyses. Retrospective analyses across the major study types are being performed to assess the value of entire studies or components of a study. For example, rat and rabbit prenatal developmental toxicity studies are mandated through FQPA and an analysis of the value of both species is being assessed in reference to its regulatory impact. Multigeneration reproductive toxicity studies have traditionally gone through two generations and the value of the second generation is being assessed for its regulatory impact and the analysis may also influence study design changes in subsequent guideline studies.

Beyond ToxRefDB and the initial anchoring to HTS and genomic data, there is a need to address data analysis and interpretation issues due to chemical metabolism and bioactivation. As observed in much of the traditional toxicity study results, species concordance is limited and the lack of concordance can be highly attributed to pharmacokinetic (PK) and pharmacodynamic (PD) differences between species (Henderson 1996). One piece of the PK/PD species differences is the capability to biotransform the

parent compound to its active metabolite. With biotransformation required for various compounds to demonstrate their toxicities, *in vitro* assays without metabolic capabilities have the potential to miss the relevant activities associated with adverse outcomes observed *in vivo*. Metabolism prediction and metabolic study data can be used to assist in identifying chemicals that require metabolic activation. However, running all potential active metabolites for even 300 chemicals through assays that do not have or have limited metabolic capability would require vast financial resources and chemical procurement may not even be possible. Since most screening programs have been limited primarily to testing parent compounds and only a few metabolites, incorporating metabolic activation into the analysis process, that is linking assay data to *in vivo* outcomes, can be performed from three different views: chemistry, biology and informatics. Using chemistry to predict potential metabolites or supply known active metabolites can help identify and filter out negative assay results possibly due to a lack of metabolic activation, but is limited in its application of further developing the predictions using only parent compound assay data. A biology-centric approach would involve comparing *in* vitro assay results across similar targets that do and do not have metabolic capacity. The approach would also use assay data, e.g., genomic data on phase I and II metabolism enzymes, to act as surrogates for understanding the metabolism of the parent compound. Other approaches for using *in* vitro screening data to predict human drug metabolism have been performed or proposed (Jolivette and Elkins 2007). The final biological piece would be to use the anchoring endpoints derived from ToxRefDB and current parent-metabolite pairs being tested within *in vitro* assays to provide reference cases. Understanding the biology will ultimately assist in interpreting the data, but may be limited in its scalability and applicability across all 300 chemicals and beyond due to the amount of

rigorous scientific inquiry for each potential case. An informatic-centered approach would involve utilizing prediction algorithms to tease out assays, with or without metabolic capacity, that provide the greatest predictability and therefore could be applied to all chemicals and endpoints. The greatest limitation with such an approach is that potentially valuable and highly informative assays results may be deemed irrelevant due to the lack of direct predictability for any given endpoint.

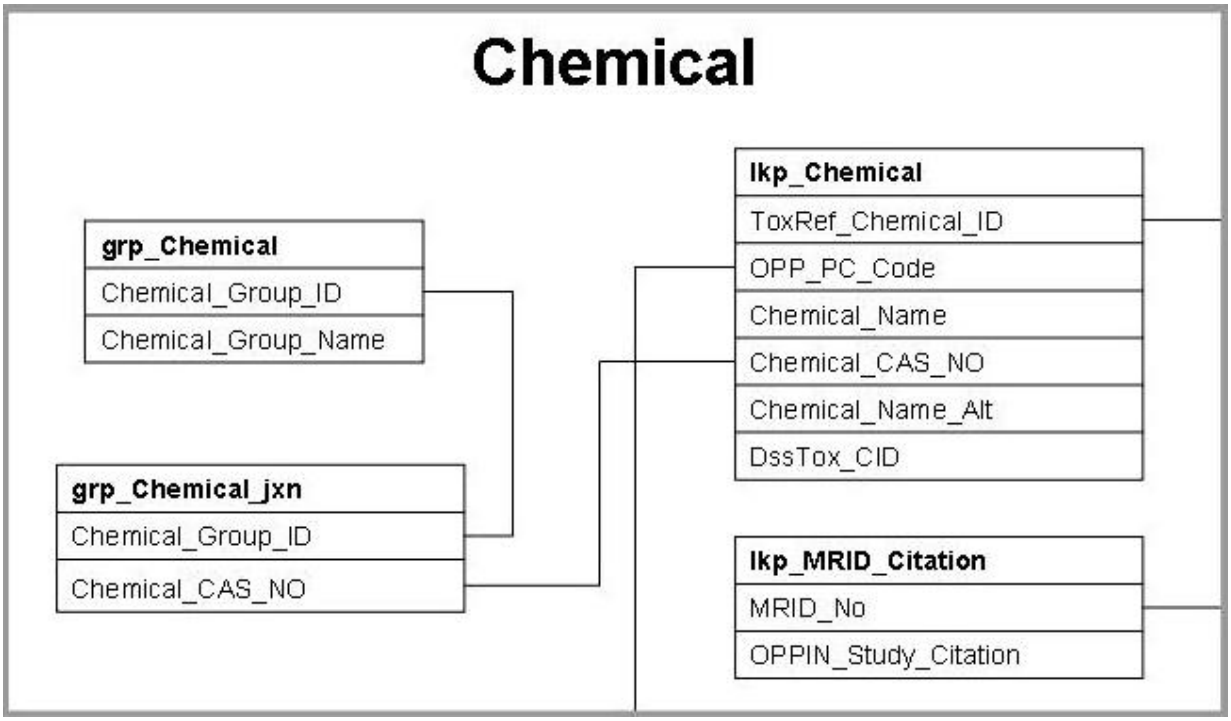
An obvious solution is to integrate the approaches and information derived from these approaches to bolster the final prediction models. Modeling and systems biology approaches integrating SAR models and biological data from HTS experiments have been proposed (Bugrim et al 2004), but the application to toxicity prediction models and high-throughput analysis has yet to be demonstrated. The testable hypothesis is that incorporating chemical and biological information into the analysis process will enable and strengthen prediction of toxicities caused by active metabolites and for which only the parent chemical was tested.

Initially, the prediction models will be used to prioritize further toxicity testing for chemicals in which little to no toxicity data exist. Beyond chemical prioritization, the vast amount of data being digitized and generated may have application to hazard and risk assessment. Biochemical, molecular, cellular and model organism data can be placed into the context of mode-of-action (MOA) and human relevancy frameworks as described in Meek et al (2003) and Dellarco and Baetcke (2005). In conjunction with detailed toxicity data from ToxRefDB, *in vitro* assay and model organism data can assist in identifying key events leading to adverse outcomes in a systematic and transparent fashion. Additionally, the diversity of cell types, both rodent and human, could be used to inform species extrapolation and human relevancy. Addressing the role of metabolism on interpreting *in vitro* assay and

extending the assay data and resulting prediction models to the risk assessment arena will move toxicology toward a more predictive and mechanistic science.

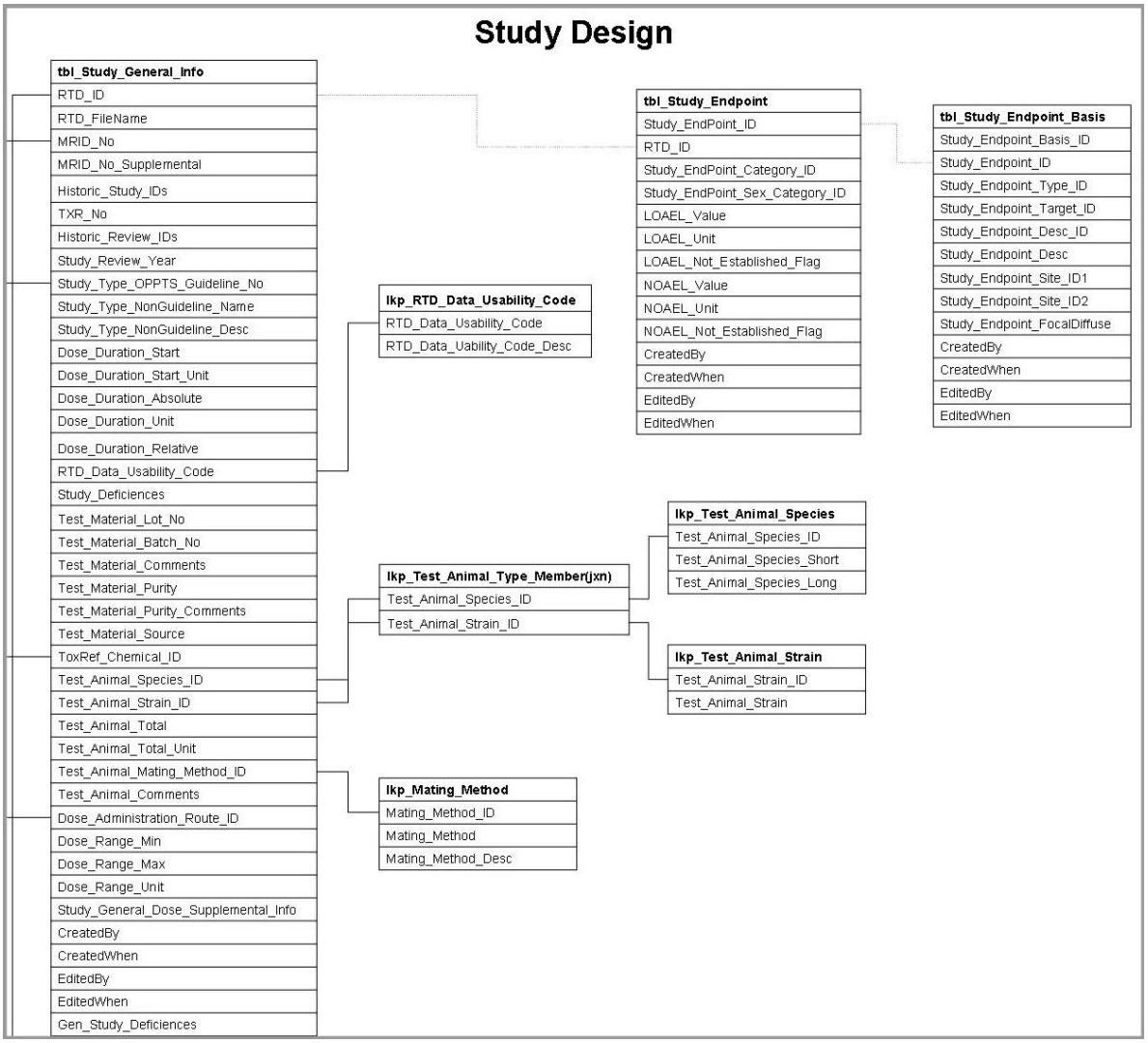
APPENDIX A

Chemical Component of Relational Data Model



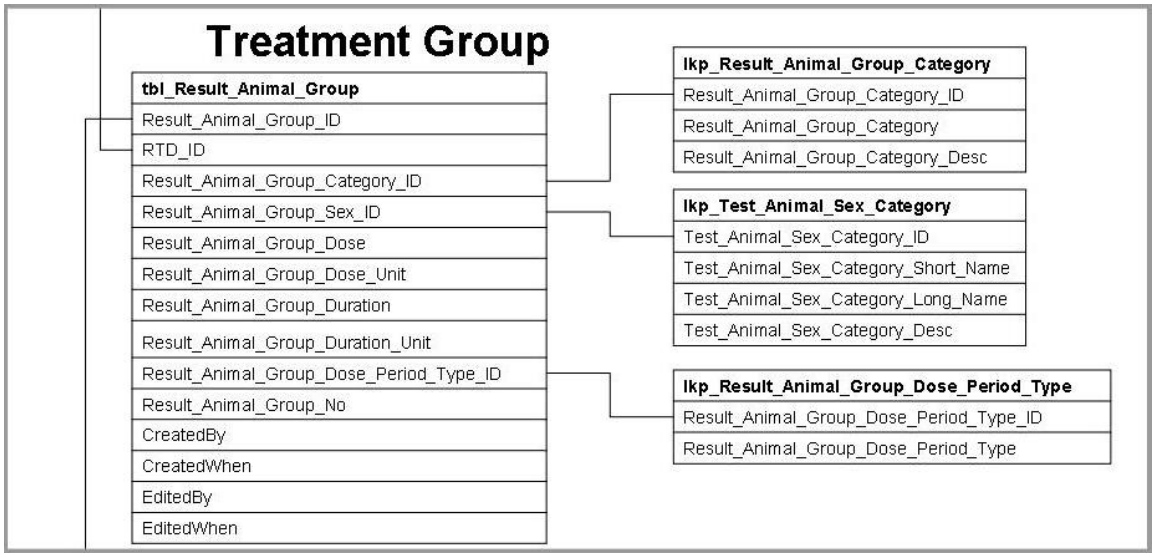
APPENDIX B

Study Design Component of Relational Data Model



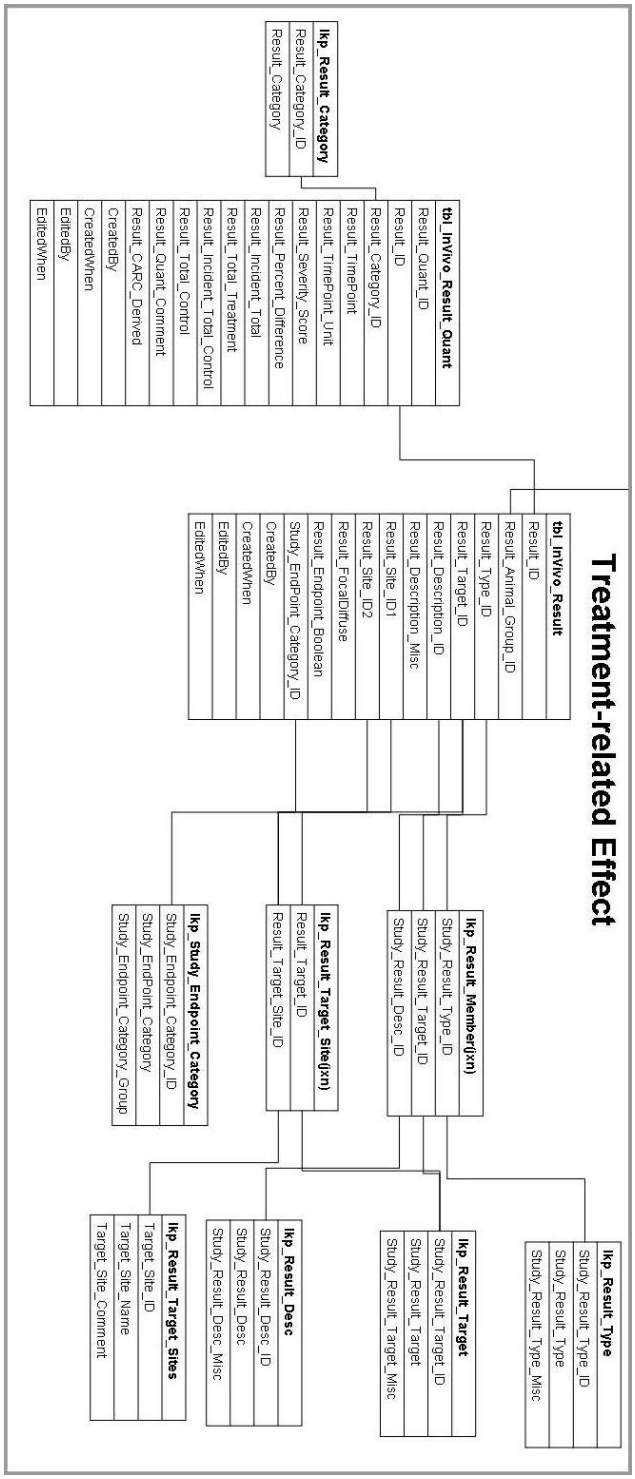
APPENDIX C

Treatment Group Component of Relational Data Model



APPENDIX D

Treatment-related Effect Component of Relational Data Model



BIBLIOGRAPHY

Allen DG, Pearse G, Haseman JK, Maronpot RR. 2004. Prediction of rodent carcinogenesis: An evaluation of prechronic liver lesions as forecasters of liver tumors in NTP carcinogencity studies. Toxicologic Pathology 32:393-401.

Apic G, Ignjatovic T, Boyer S, Russell RB. 2005. Illuminating drug discovery with biological pathways. FEBS Lett 579:1872-1877.

Benigni R. 1991. QSAR prediction of rodent carcinogenicity for a set of chemicals currently bioassayed by the US National Toxicology Program. Mutagenesis 6:423-425.

Benigni R., Zito R. 2004. The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. Mutation Research 566:49-63.

Bitsch A, Jacobi S, Melber C, Wahnschaffe U, Simetska N, Mangelsdorf I. 2006. REPDOSE: A database on repeated dose toxicity studies of commercial chemicals – A multifunctional tool. Regulatory Toxicology and Pharmacology 26:202-210.

Bugrim A, Nikolskaya T, Nikolsky Y. 2004. Early prediction of drug metabolism and toxicity: systems biology approach and modeling. Drug Discovery Today 9:127-135.

Collins FS, Gray GM, Bucher JR. 2008. Transforming environmental health protection. Science 319:906-907.

Dellarco VL, Baetcke K. 2005. A risk assessment perspective: application of mode of action and human relevance frameworks to the analysis of rodent tumor data. Toxicol Sci 81:1-3.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95:5-12.

Felsot AS. 2002. WEB resources for pesticide toxicology, environmental chemistry, and policy: a utilitarian perspective. Toxicology 173(1-2): 153-166.

Fielden MR, Brennan R, Gollub J. 2007. A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. Toxicol Sci 99:90-100.

Fielden MR, Eynon BP, Natsoulis G, Jarnagin K, Banas D, Kolaja KL. 2005. A gene expression signature that predicts the future onset of drug-induced renal tubular toxicity. Toxicol Pathol 33:675-683.

Fielden MR, Matthews JB, Fertuck KC, Halgren RG, Zacharewski TR. 2002. In silico approaches to mechanistic and predictive toxicology: an introduction to bioinformatics for toxicologists. Crit Rev Toxicol 32:67-112.

Fliri AF, Loging WT, Thadeio PF, Volkmann RA. 2005. Analysis of drug-induced effect patterns to link structure and side effect of medicines. Nature Chem Biol 1(7):389-397.

Gold LS, Manley NB, Slone TH, Ward JM. 2001. Compendium of Chemical Carcinogens by Target Organ: Results of Chronic Bioassays in Rats, Mice, Hamsters, Dogs, and Monkeys. Toxicol Pathology 29:639-652.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell 100:57-70.

Henderson RF. 1996. Species differences in the metabolism of benzene. Envir Health Perspectives. 104:1473-1175.

Ihaka R, Gentleman R. 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics. 5: 299-314.

Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, Auld DS. 2007. High-throughput screening assays for the identification of chemical probes. Nature Chemical Biology 3:466-479.

Janer G, Hakkert BC, Slob W, Vermeire T, Piersma AH. 2007a. A retrospective analysis of the two-generation study: What is the added value of the second generation? Reproductive Toxicology 24:97-102.

Janer G, Slob W, Hakkert BC, Vermeire T, Piersma AH. 2007b. A retrospective analysis of developmental toxicity studies in rat and rabbit: What is the added value of the rabbit as an additional test species? Regulatory Toxicology and Pharmacology 50:206-217.

Jolivette LJ, Elkins S. 2007. Methods for predicting human drug metabolism. Advances in Clinical Chemistry 43: 131-176.

Judson R, Dix D J, Houck K, Richard A M, Martin MT, Kavlock RJ, Dellarco V, Holderman T, Tan S, Carpenter T, Smith E. In press. The Toxicity Data Landscape for Environmental Chemicals.

Judson RS, Richard AM, Dix DJ, Houck K, Elloumi F, Martin MT, et al. 2008. ACToR – Aggregated Computational Toxicology Resource. Toxicology and applied pharmacology Submitted.

Junghans TB, Sevin IF, Ionin B, Seifried H. 2002. Cancer information resources: digital and online sources. Toxicology 173(1-2): 13-34.

Kavlock RJ, Ankley G, Blancato J, Breen M, Conolly R, Dix D, Houck K, Hubal E, Judson R, Rabinowitz J, Richard A, Setzer W, Shah I, Villeneuve D, Weber E. 1997. Computational Toxicology – A State of the Science Mini Review. Toxicol. Sci. doi:10.1093/toxsci/kfm297

Kavlock RJ, Dix DJ, Houck KA, Judson RS, Martin MT, Richard AM. In press. ToxCast™: Developing predictive signatures for chemical toxicity. Proceedings from the 6th World Congress on Alternatives & Animal Use in the Life Sciences.

Martin MT, Brennan RJ, Hu W, Ayanoglu E, Lau C, Ren H, Wood CR, Corton JC, Kavlock RJ, Dix DJ. 2007. Toxicogenomic study of triazole fungicides and perfluoroalkyl acids in rat livers predicts toxicity and categorizes chemicals based on mechanisms of toxicity. Toxicol Sci 97(2):595-613.

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF. 2005. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regulatory Toxicology and Pharmacology 44:83-96.

Mayne JT, Ku WW, Kennedy SP. 2006. Informed toxicity assessment in drug discovery: systems-based toxicology. Curr Opin Drug Discov Devel 9:75-83.

Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, Longfellow DG, Pastoor T, Seed J, Patton DE. 2003. A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol 33:591-653.

NTP (National Toxicology Program). 2004. A National Toxicology Program for the 21st Century: A Roadmap for the Future. Available: http://ntp.niehs.nih.gov/go/vision [accessed 14 March 2008].

NTP (National Toxicology Program). 2007. Pathology Code Tables. Available: http://jones.niehs.nih.gov:8080/pct/ [accessed 7 February 2008].

NRC (National Research Council). 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. The National Academies Press, Washington, DC, 196pp.

Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, et al. 2000. Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals, Regulatory Toxicology and Pharmacology 32:56-67.

Patterson J, Hakkinen PJ, Wullenweber AE. 2002. Human health risk assessment: selected Internet and world wide web resources. Toxicology 173(1-2): 123-143.

Paules R. 2003. Phenotypic anchoring: Linking cause and effect. Environmental Health Perspectives. 111(6):A339.

Powell CL, Kosyk O, Ross PK, Schoonhoven R, Boysen G, Swenberg JA, Heinloth AN, Boorman GA, Cunningham ML, Paules RS, Rusyn I. 2006. Phenotypic anchoring of acetaminophen-induced oxidative stress with gene expression profiles in rat liver. Toxicol Sci 93:213-22.

Richard A, Williams C. 2003. Public Sources of Mutagenicity and Carcinogenicity Data: Use in Structure-Activity Relationship Models. In: QSARS of Mutagens and Carcinogens (Benigni R, ed). New York: CRC Press, 145-173.

Richard A. 2004. DSSTox web site launch: Improving public access to databases for building structure-toxicity prediction models. Preclinica 2:103-108.

Roos DS. 2001. Computation Biology: Bioinformatics—Trying to swim in a sea of data. Science 291:1260-1261.

Russom CL. 2002. Mining environmental toxicology information: web resources. Toxicology 173(1-2): 75-88.

U.S. EPA (U.S. Environmental Protection Agency). 1996. OPPTS Harmonized Test Guidelines Website. Available: http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/ [accessed 10 January 2008].

U.S. EPA (U.S. Environmental Protection Agency). 1997. Integrated Risk Information System (IRIS) Website. Available: http://cfpub.epa.gov/ncea/iris/index.cfm [accessed 18 February 2008].

U.S. EPA (U.S. Environmental Protection Agency). 2007. Distributed Structure-Searchable Toxicity Database Website. Available: http://www.epa.gov/ncct/dsstox/ [accessed 14 February 2008].

Waters M, Boorman G, Bushel P, Cunningham M, Irwin R, Merrick A, et al. 2003. Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. EHP Toxicogenomics 111(1T): 15-28.

Winter CK. 2002. Electronic information resources for food toxicology. Toxicology 173(1-2): 89-96.

Wolf DC, Mann PC. 2005. Confounders in interpreting pathology for safety and risk assessment. Toxicol Applied Pharmacology 202:302-308.

Wolfgang GH, Johnson DE. 2002. Web resources for drug toxicity. Toxicology 173(1-2): 67-74.

Yang C, Benz RD, Cheeseman MA 2006a. Landscape of current toxicity databases and database standards. Curr Opin Drug Discov Devel 9:124-133.

Yang C, Richard AM, Cross KP 2006b. The art of data mining the minefields of toxicity databases to link chemistry to biology. Curr Comput-Aided Drug Dis. 2:135-150.

Young RR. 2002. Genetic toxicology: web resources. Toxicology 173(1-2): 103-121.

Zhu H, Rusyn I, Richard A, Tropsha A. 2008. The use of cell viability assay data improves the prediction accuracy of conventional quantitative structure activity relationship models of animal carcinogenicity. Env Health Perspectives doi:10.1289/ehp.10573.